

(19) 日本国特許庁 (J P)

(12) 特 許 公 報 (B 2)

(11) 特許番号

第2656586号

(45) 発行日 平成 9 年 (1997) 9 月 24 日

(24) 登録日 平成 9 年 (1997) 5 月 30 日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00			G 1 0 L 3/00	B
	5 1 3			5 1 3 A
9/00			9/00	D
9/12			9/12	D

請求項の数 5 (全 10 頁)

(21) 出願番号 特願昭63-300479
(22) 出願日 昭和63年(1988)11月30日
(65) 公開番号 特開平2-148099
(43) 公開日 平成 2 年 (1990) 6 月 6 日

(73) 特許権者 999999999
株式会社日立製作所
東京都千代田区神田駿河台 4 丁目 6 番地
(72) 発明者 村松 隆二郎
神奈川県横浜市戸塚区戸塚町216番地
株式会社日立製作所戸塚工場内
(74) 代理人 弁理士 小川 勝男 (外 1 名)

審査官 新宮 佳典

(56) 参考文献 特開 昭63-285595 (J P, A)
特開 昭56-78899 (J P, A)
特開 昭63-241600 (J P, A)
特開 昭63-259596 (J P, A)
特開 昭62-247798 (J P, A)
特開 昭63-220295 (J P, A)
特開 昭60-211500 (J P, A)

(54) 【発明の名称】 音声検出方法及びその装置

1

(57) 【特許請求の範囲】

【請求項 1】 フレーム毎に音声の有無を識別する音声検出方式において、

上記各フレームを複数のサブフレームに分割し、上記サブフレーム毎の音声情報を検出する第 1 のステップと；

上記音声情報の検出結果に重み付けを行ない、1 フレームにわたって総和を求める第 2 のステップと；

上記重み付けを行なった出力結果に基づき、音声の有無を判別する第 3 のステップと；

からなる音声検出方法。

【請求項 2】 各フレームの前半部分のサブフレームより、各フレームの後半部分のサブフレームに重み付けを行なつて、語頭切れをなくすようになしたことを特徴とする請求項 1 記載の音声検出方法。

【請求項 3】 各フレームの後半部分のサブフレームによ

2

り、各フレームの前半部分のサブフレームに重み付けを行なつて、語尾切れをなくすようになしたことを特徴とする請求項 1 記載の音声検出方法。

【請求項 4】 フレーム毎に音声の有無を識別する音声検出方式において、

上記各フレームを複数のサブフレームに分割し、上記サブフレーム毎の音声情報を検出する第 1 のステップと；

上記各フレームの前半部分のサブフレームの音声情報の検出結果より、上記各フレームの後半部分のサブフレームの音声情報の検出結果に重み付けを行い、1 フレームにわたって総和を求める第 2 のステップと；

上記各フレームの後半部分のサブフレームの音声情報の検出結果より、上記各フレームの前半部分のサブフレームの音声情報の検出結果に重み付けを行い、1 フレームにわたって総和を求める第 3 のステップと；

10

を備え、上記第2のステップによつて音声の検出出力が得られない時に、上記第3のステップで音声の検出を行なうようにしたことを特徴とする音声検出方法。

【請求項5】フレーム毎に音声の有無を識別する音声検出方式において、

上記各フレームを複数のサブフレームに分割し、上記サブフレーム毎の音声情報を検出する音声情報検出手段と；

上記音声情報検出手段の検出結果に重み付けを行ない、1フレームにわたつて総和を求める重み付け総和手段と；

上記重み付け総和手段からの出力結果に基づき、音声の有無を判別する判別手段と；

を備えたことを特徴とする音声検出装置。

【発明の詳細な説明】

〔産業上の利用分野〕

本発明は、情報通信システムのデジタル・スピーチ・インタポレーション (Digital Speech Interpolation, 以下、DSIと称す) において、特にフレーム処理を行なう高能率音声符号化装置に好適な音声検出方式に関する。

〔従来の技術〕

従来、音声検出器については、NTCレコード10、6-1 (1976年) (National Telecommunications Conference Record 10.6-1, Dec. 1976) において論じられている。

すなわち、従来の音声検出器では、入力音声を低域通過フィルタに通してエンベロープ成分を抽出するかまたは入力音声よりパワーを直接演算により求めるかした後、固定閾値と比較して有音無音の識別を連続的に行なつて来た。

そして、この音声検出出力を用いて、無音区間を抑圧して、回線の有効活用をはかるDSIシステム等の制御を行なつて来た。

〔発明が解決しようとする課題〕

一般的に、フレーム単位 (通常、フレーム周期は20ms) に音声を切り出してまとめて情報圧縮を行なう高能率符号化方式により、音声情報を伝達するDSIシステムにおいては、音声検出出力も、フレーム周期毎に出力され、該当フレームが有音フレームであるか、無音フレームであるかの識別に利用される。

しかしながら、このようなDSIシステムにおいて、従来技術である音声検出方式を適用した場合、特定フレーム内に存在する入力音声のエンベロープまたはパワーにより有音無音の識別を行なうため、フレームの後半で無音から有音に立ち上がる音声や、フレームの前半で有音から無音に変化する音声の場合該当フレームは、無音フレームとして識別され易くなる。そして、このような現象はシステムを通して通話した場合、それぞれ語頭切れ、語尾切れにつながり、正しく会話内容が伝わらない

という不具合を生ずる。また、前述の不具合を回避するため単に検出閾値を下げたのでは、微弱な雑音を有音と識別し易くなり、システムの回線利用効果の低下につながる。

〔課題を解決するための手段〕

本発明は、フレーム毎に音声の有無を識別する音声検出方式において、

上記各フレームを複数のサブフレームに分割し、上記サブフレーム毎の音声情報を検出する第1のステップと；

上記音声情報の検出結果に重み付けを行ない、1フレームにわたつて総和を求める第2のステップと；

上記重み付けを行なつた出力結果に基づき、音声の有無を判別する第3のステップと；

からなることを特徴とする。

〔作 用〕

本発明の音声検出器によれば、無音から有音へ変化時には、乗じられる重みがフレームの後半になるほど大きいために、音声が入力し始める時点がフレームの前半にある場合と、フレームの後半にある場合とで、フレーム単位の音声情報の値は、従来技術に比べて差が小さくどちらにも有音と判定され易くなる。また、有音から無音への変化時には、乗じられる重みがフレームの後半になるほど小さいために、音声がなくなる時点がフレームの前半にある場合と、フレームの後半にある場合とで、フレーム単位の音声情報の値は、やはり従来技術と比べて差が小さくどちらにも有音と判定され易くなる。これにより、従来の技術で問題となっていた語頭切れや語尾切れを少なくすることができる。なお、微少雑音に対しては、サブフレーム毎の検出閾値が雑音レベルを下まわらない様に設定しておくことにより、誤つて有音と識別することを防止することができる。

〔実施例〕

以下、本発明の実施例を図面を参照して詳細に説明する。

第3図は、長距離電話伝送において、回線の有効活用をはかるDSIシステムの構成図である。

同図において、20及び70は、構内交換機 (以下、PBXと称す)、11～1n及び81～8nは、それぞれPBX20、PBX70の内線電話端末、5及び6はデジタル回線、1は、PBX20の内線電話端末10～1nからの音声データが有音か無音かを検出し、有音である音声データのみ多重化してデジタル回線5へ送信するDSI送信装置、2は、デジタル回線5からの多重化された音声信号を受信して、分配制御を行ない、PBX70の内線電話端末80～8nへ出力させるDSI受信装置、3は、PBX70の内線電話端末80～8nからの音声データが有音か無音かを検出し、有音である音声データのみを多重化してデジタル回線6に送信するDSI送信装置、4は、デジタル回線6からの多重化されたデータを受信して、分配制御を行ないPBX20の内端

電話端末10～1nへ出力させるDSI受信装置である。

まず、内線電話端末11～1nより内線電話端末81～8nへの音声データの伝送に関して、本システムのDSI送信装置1の動作を第4図を参照し説明する。

内線電話端末11～1nより入力した音声はPBX20を通った後、DSI送信装置1に入力される。DSI送信装置1に入力された音声は、高能率音声符号器31～3nにより20msフレーム単位で情報圧縮されると同時に、音声検出器41～4nによつて音声の有無を20ms毎に検出される。音声検出器41～4nからの音声検出出力は割当制御部51に入力され、音声信号接続部50を制御すると共に、割当信号発生部52へ情報を伝達する。音声信号接続部50では、n本の高能率音声符号器31～3nからの出力のうち有音部の音声符号化データ出力のみを多重化し回線多重部53へ送出する。回線多重部53は、更に上記有音部の音声符号化データと割当信号を多重化した多重化データをデジタル回線5へ出力する。

次に、本システムのDSI受信装置2の動作について第5図を参照し説明する。

まず、デジタル回線5からの多重化データは、多重分離部54によつて上記音声符号化データと上記割当信号*

$$P_{sk} = \frac{1}{32} \sum_{i=1}^{32} X_i^2 \dots\dots\dots (1)$$

X_i : フィルタ出力

k : サブフレーム番号

次に、サブフレーム内パワー演算部4103及び4105で
は、各々高レベル検出閾値4104、低レベル検出閾値4106
との比較が次の様に行なわれる。

検出閾値を T_{hj} で表せば、

$P_{sk} \geq T_{hj}$ のとき $D_{jk} = 1$

$P_{sk} < T_{hj}$ のとき $D_{jk} = 0$

$$Z_{sk} = \sum_{i=1}^{32} \text{sgn}(X_{i-1} \cdot X_i) \dots\dots\dots (2)$$

ここで

$$\text{sgn}(X_{i-1} \cdot X_i) = \begin{cases} 0 : X_{i-1} \text{ と } X_i \text{ が同符号} \\ 1 : X_{i-1} \text{ と } X_i \text{ が異符号} \end{cases}$$

である。

そして、ゼロクロス数検出閾値4108との比較が同様に次の様に行なわれる。

$Z_{sk} \geq T_{h3}$ のとき $D_{3k} = 1$

$Z_{sk} < T_{h3}$ のとき $D_{3k} = 0$

* とに分離され、音声符号化データを音声信号接続部55へ、割当信号を割当信号受信部56へそれぞれ入力する。次に、分配制御部57は、割当信号受信部56からの出力をもとに音声信号接続部55の制御を行い、音声信号接続部55は、音声符号化データの分配制御を行なう。そして、分配制御された音声符号化データは、音声復号器61～6nによつて音声データに変換され、PBX70を介して内線電話端末81～8nへ出力される。

また、同様にしてPBX70の内線電話端末81～8nにより入力した音声は、DSI送信装置3で多重化され、DSI受信装置4によつて再生されて、PBX20の内線電話端末11～1nへ出力される。

次に、DSI送信装置1及び3の音声検出器(第4図, 41～4n)の動作を第1図, 第2図により説明する。

第1図において、音声入力4101は高域通過フィルタにより直流成分が除去された後、4ms長のサブフレーム毎に動作するサブフレーム内パワー構算部4103及び4105とサブフレーム内ゼロクロス数計数部4107に入力される。

音声入力は8KHzでサンプリングされているためサブフレーム内パワー演算部4103及び4105では次式(1)によりパワーを算出する。

※ (j = 1, 2)

によりパワー検出を行なう。

また、同時にサブフレーム内ゼロクロス数計数部4107によつて連続する2サンプルのフィルタ出力の極性ビットが異符号となる数、すなわちゼロクロス数の計数を次の演算(2)によつてサブフレーム毎に行なわれる。

★ 次に、サブフレーム内パワー演算部4103, 4105及びサブフレーム内ゼロクロス数計数部4107の出力 D_{1k} , D_{2k} 及び D_{3k} は、それぞれ増加関数重み付け総和部4109, 4111, 4113に入力され、次の演算(3)が行なわれる。

★

$$D_j = \sum_{k=1}^5 k \cdot D_{jk} \quad \dots\dots\dots (3)$$

$$(j = 1, 2, 3)$$

また、減少関数重み付け総和部4112では、次の演算 * * (4) が行なわれる。

$$D_4 = \sum_{k=1}^5 (6 - k) \cdot D_{4k} \quad \dots\dots\dots (4)$$

なお、上述した増加関数重み付け演算の様子を第6図に示し、また減少関数重み付け演算の様子を第7図に示す。

以上により求められた D_1, D_2, D_3, D_4 は、フレーム単位の重み付け総和であり、整数値となる。また、これらを再び固定閾値4114~4117と、フレーム毎の検出部4118, 4120, 4121, 4122において次の様に比較する。

※ $D_j \geq F_{mj}$ のとき $L_j = 1$
 $D_j < F_{mj}$ のとき $L_j = 0$

($j = 1, 2, 3, 4$)

ここで、 F_{mj} は固定閾値4114~4117の閾値を表わしている。

また、これらとは別にフレーム内パワー演算部4110によつて、フレーム内パワーを次式(5)により求める。

$$P_n = \sum_{k=1}^5 P_{sk} \quad \dots\dots\dots (5)$$

(n : フレーム番号)

次にフレーム間パワー増分検出器4119にて、下記判定★ ★を行なう。

$$P_n \geq \alpha P_{n-1} \quad \text{のとき} \quad L_s = 1$$

$$\alpha > 1$$

$$P_n < \alpha P_{n-1} \quad \text{のとき} \quad L_s = 0$$

これは、現フレームのパワーが前フレームのパワーの2倍以上かどうかを判定するものである。

上述の動作によつて得られた $L_1 \sim L_5$ をもとに、論理判定部4123は、有音、無音の判定を行なう。

以下、論理判定部4123の動作を、第2図に従い説明する。

論理判定部4123は、まず、検出部4118の出力 L_1 が“1”(step1) かまたは、検出部4122の出力 L_5 が“1”(step2) の場合には、有音と判定し、有音フラグに“1”をセットし(step9)、ハング・オーバー・タイマをセットして(step10)、有音フラグの出力を行なう(step11)。また、検出部4120の出力 L_2 およびパワー増分検出器の出力 L_3 が同時に“1”の場合も(step3, step4)、同様に有音と判定して、有音フラグを“1”にセットし(step9)、ハング・オーバー・タイマをセットして(step10)、有音フラグの出力(“1”)を音声検出力4124へ行なう(step11)。

次に、検出部4121の出力 L_4 が“0”となつた場合には(step5)、論理判定部4123の中にあるハング・オーバー・タイマを起動し(step6, step7)、タイマー設定時間経過後有音フラグを“0”にセットして(step8)、有

☆ 音フラグの出力(“0”)を音声検出力4124へ行なう(step11)。

30 [発明の効果]

本発明によれば、以下に記載される効果が得られる。

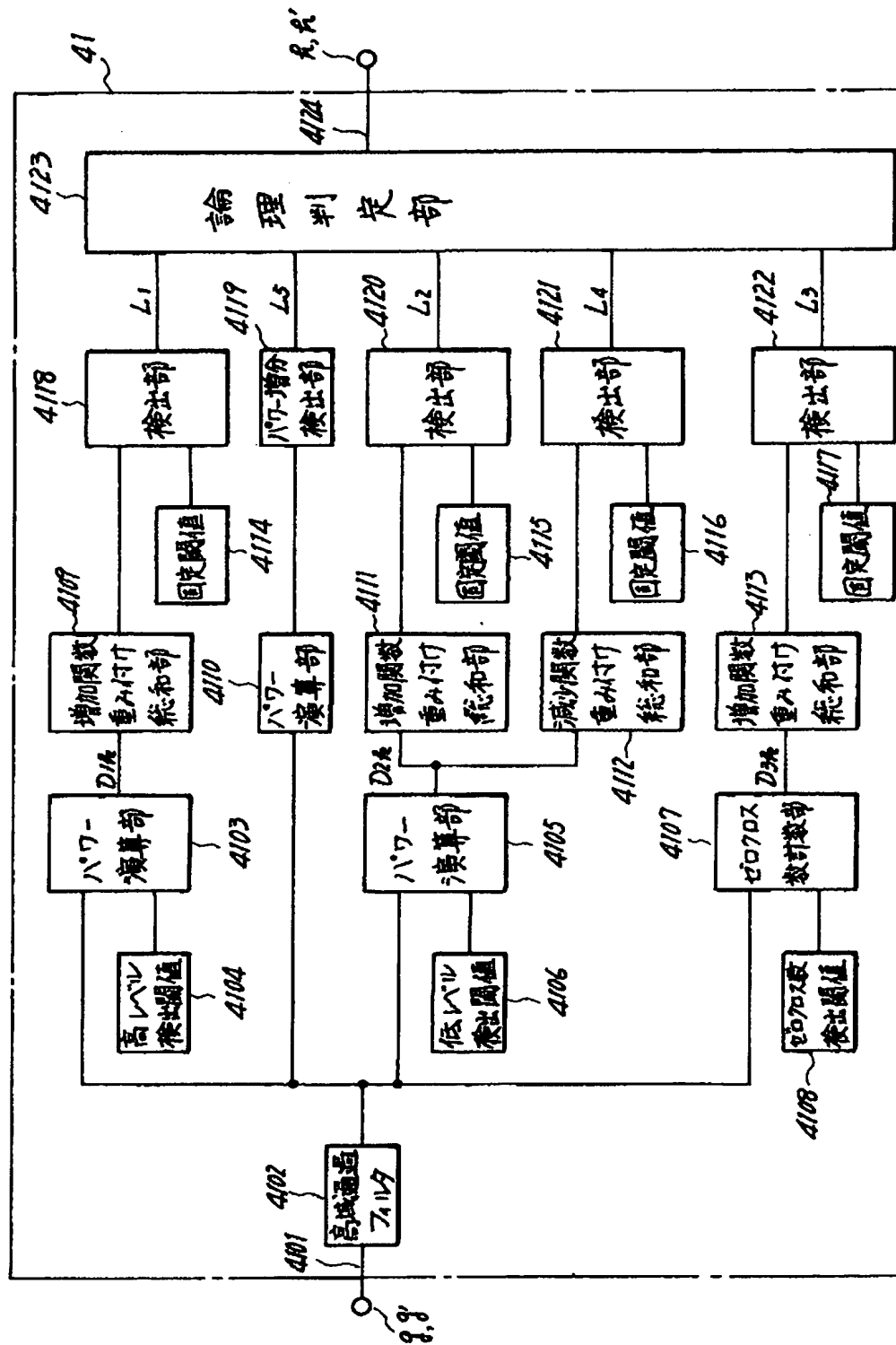
(1) 音声開始時には、フレーム後半のサブフレームの音声検出結果に大きな重みをかけることにより語頭切れの少ない音声検出が行なえる。

(2) 音声終了時には、フレームの前半のサブフレームの音声検出結果に大きな重みをかけることにより語尾切れの少ない音声検出が行なえる。

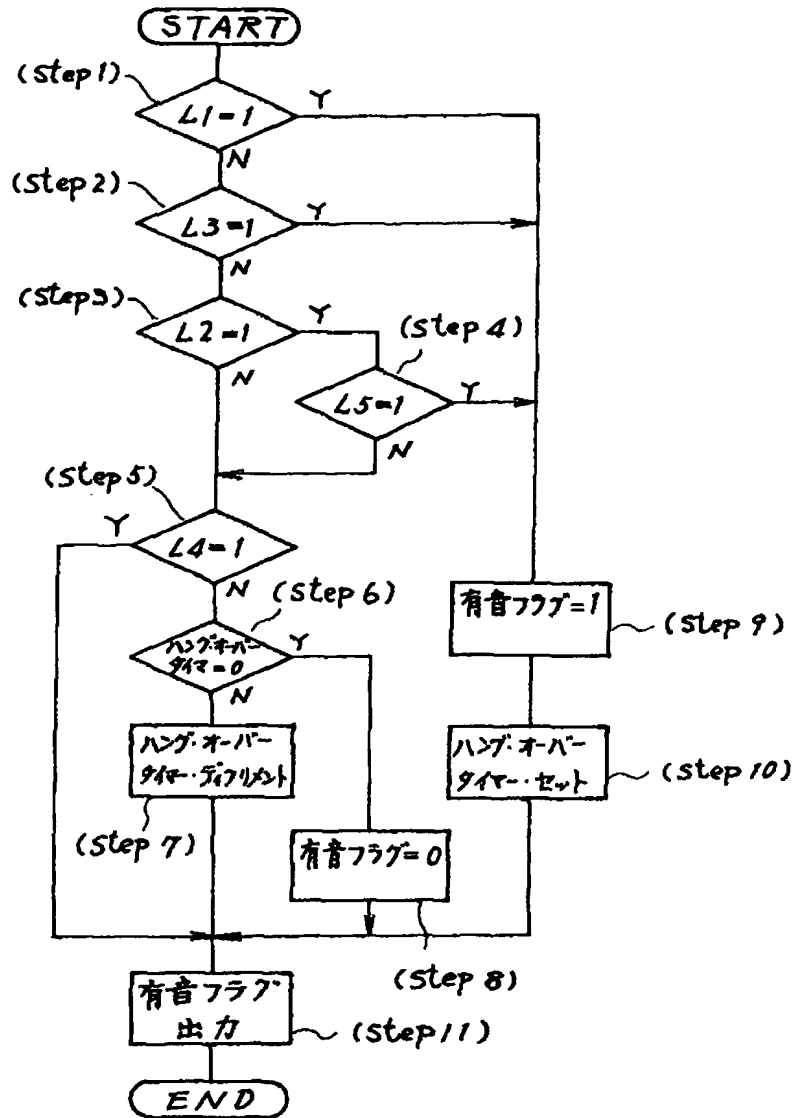
【図面の簡単な説明】

第1図は、音声検出器のブロック構成図、第2図は、論理判定部の論理判定フローチャート図、第3図は、DSIシステム構成図、第4図は、DSI送信装置のブロック構成図、第5図は、DSI受信装置のブロック構成図、第6図は、フレームの後半で立ち上がる音声を、有音と判定する様子を示す図、第7図は、フレームの前半で立ち下がる音声を、有音と判定する様子を示す図である。
 1, 3……DSI送信装置、2, 4……DSI受信装置、20, 70……構内交換機、11~1n, 81~8n……内線電話端末、41……音声検出器。

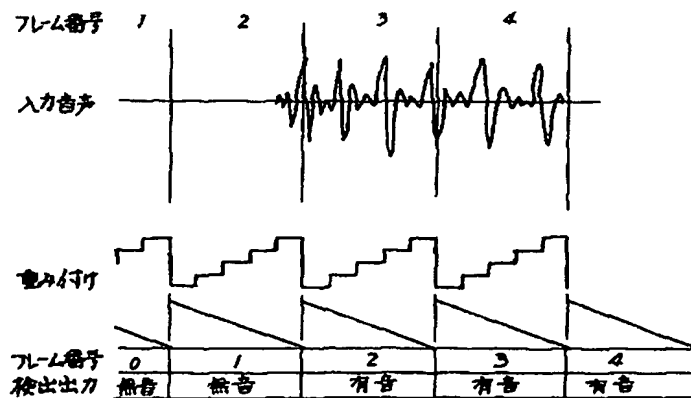
【第 1 図】



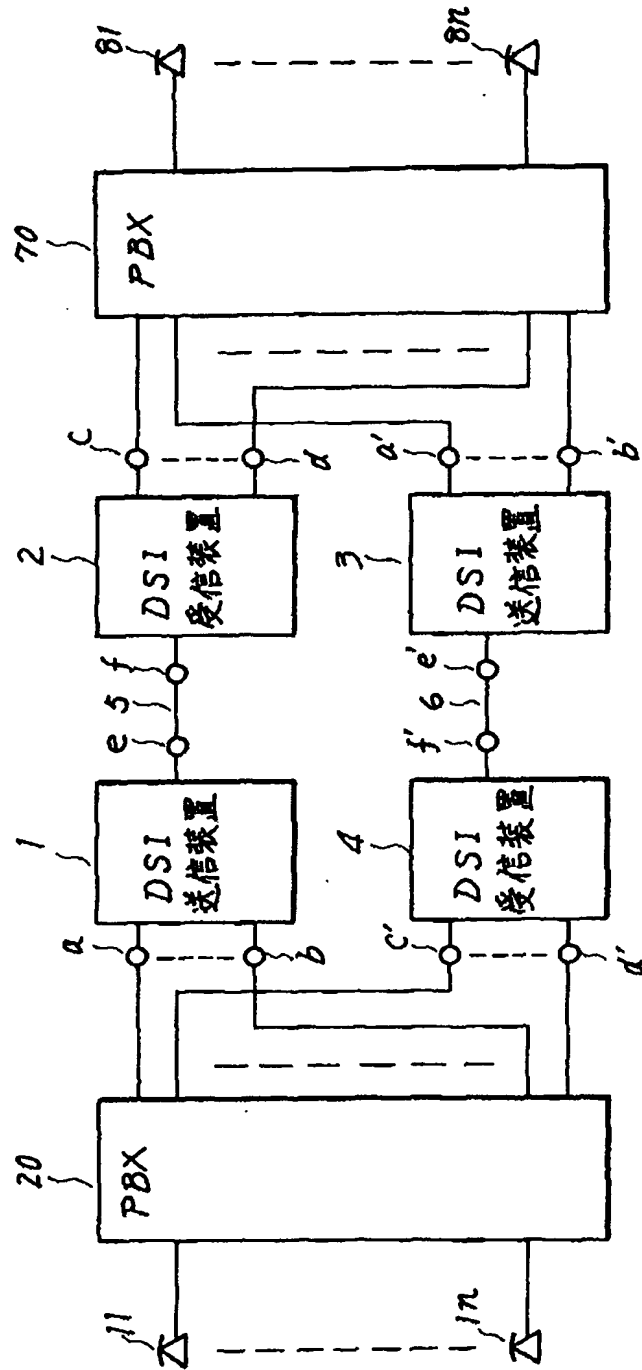
【第2図】



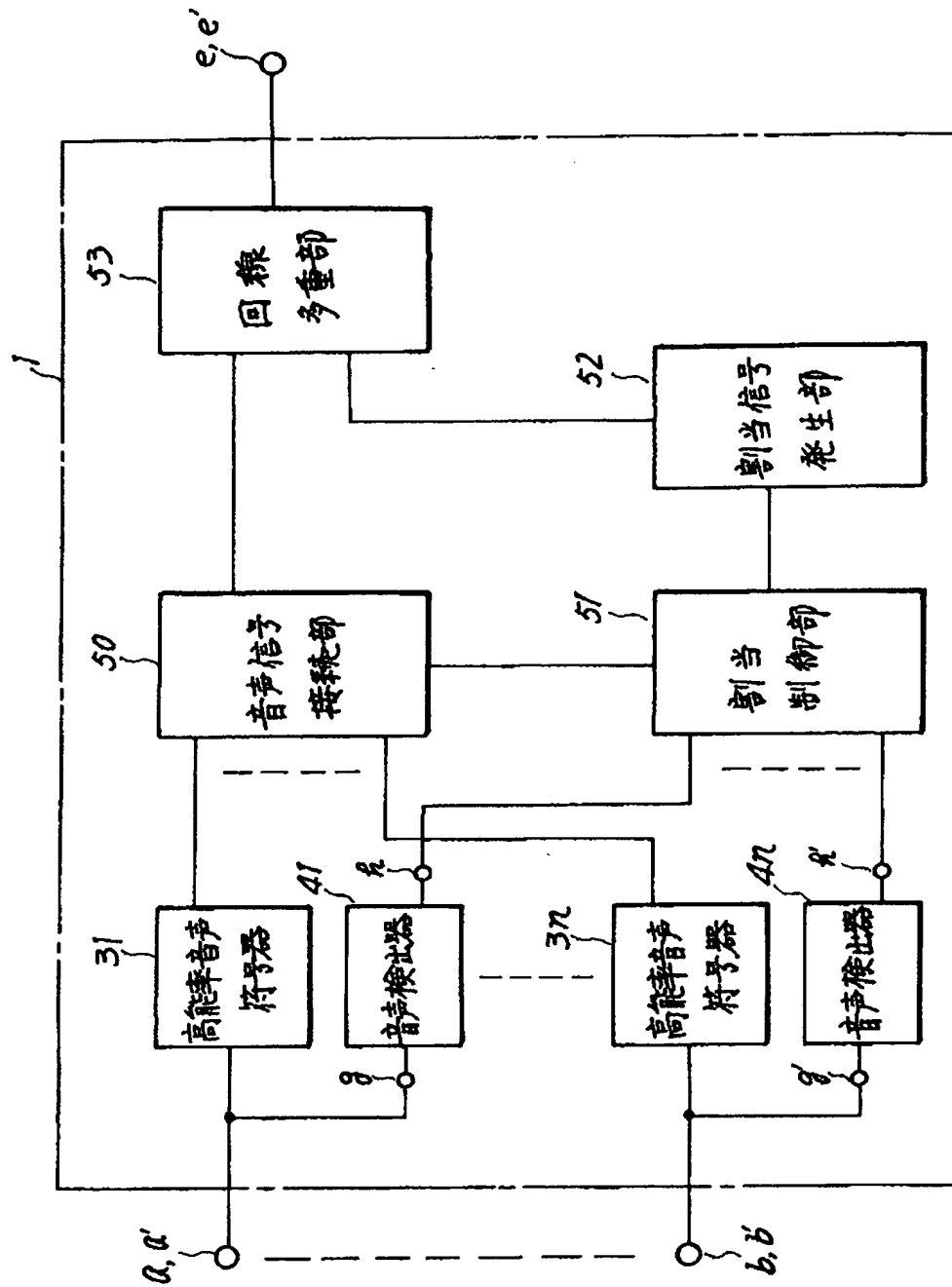
【第6図】



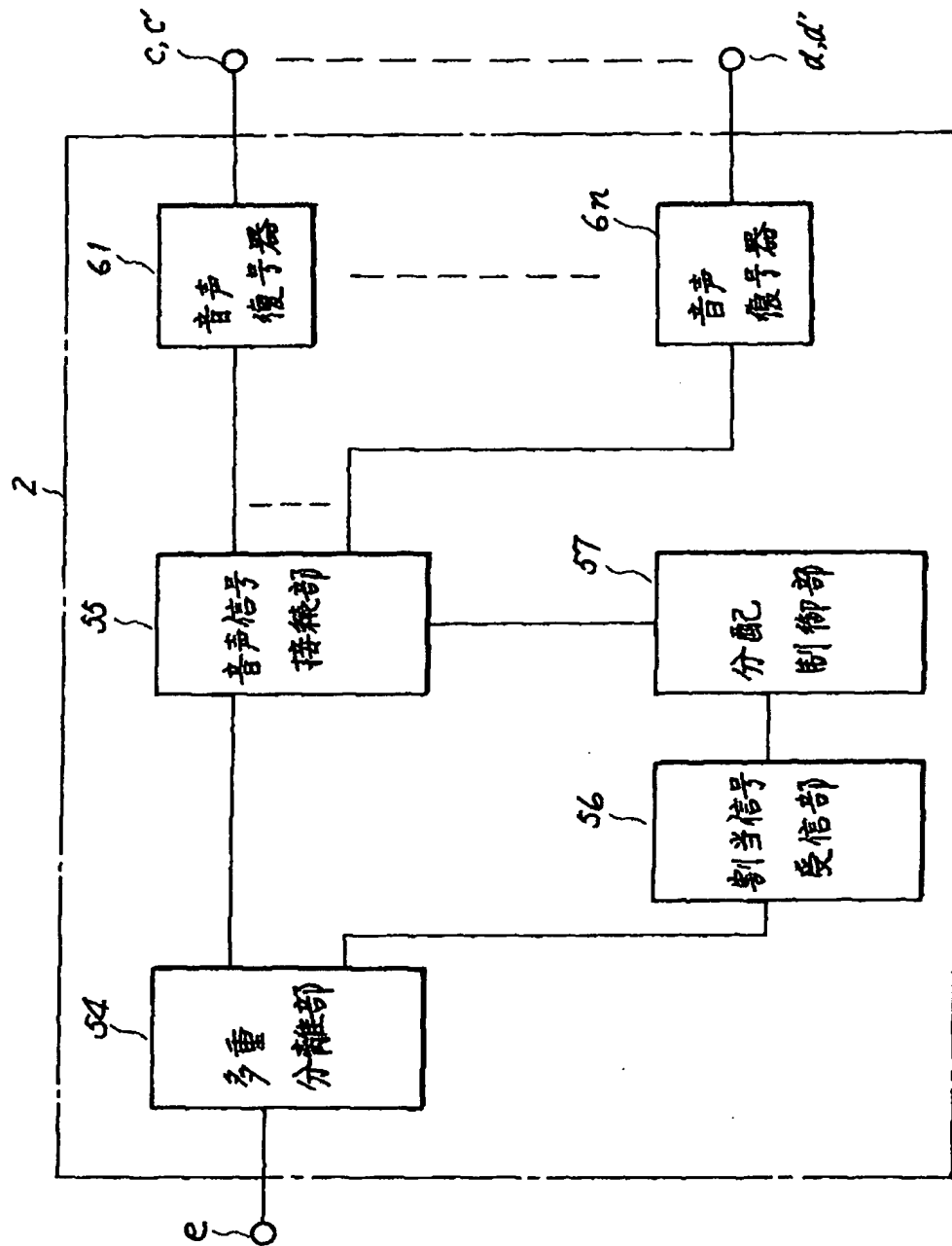
【第3図】



【第 4 図】



【第5図】



【第7図】

